AMENDMENTS TO THE SPECIFICATION

Please amend the Specification at page 8, line 32 in the following manner:

Figure 18 shows Table 2, Selected Seed-Specific Genes. The selected ESTs and their predicted protein sequences were blasted against protein and DNA sequence databases of NCBI, to identify a possible function of each gene and its corresponding Arabidopsis genome sequence.

Please amend the Specification at page 9, line 1 in the following manner:

Figure 19 shows Table 3, "Primers for PCR Amplification of 12 Promoter Regions." Provided in Table 3 is the name, sequence, position, REs, T(°C), Length 1, Length 2. Position is the distance from the best position (for reverse primers, it is ATG). REs is the included restriction enzyme site. T(°C) is the annealing temperature. Length 1 is the length of the sequences existing in the genomic sequences. Length 2 is the full length.

Please amend the Specification at page 34, line 18 through page 35, line 8 in the following manner:

The discovery of the seed-specific promoter regions is a result of a genome wide analysis of gene expression in developing seeds of *Arabidopsis thaliana*. This discovery process can be divided into several stages; the first is the isolation and analysis of expressed sequence tags (ESTs) and their associated cDNAs from developing seeds. The next stage is a microarray analysis of a selected subset of the ESTs; this analysis is used to broadly analyze the expression of several thousand genes during seed development, to identify tissue-specific expression patterns, and to identify certain genes for further analysis. The next stage is to select a second EST subset within the initial subset, where the second subset comprises ESTs which are identified as highly expressed and seed-specific. In stage three, genome sequences which match the ESTs noted above are identified from the *Arabidopsis* genome. The flanking sequences of these genes are analyzed by software programs, such as GeneScan, GeneStart, and Genefinder (which are publically available gene prediction programs ~~accessible through the site arabidopsis.org/geneid.html~~) to predict the associated promoter regions. Next, a subset of

the gene promoter regions is identified and characterized. Identification is based generally on: a) comparison to protein sequences, if available; b) high probability of ATG prediction; and c) high probability of gene prediction, and characterized. Characterization includes determining the effectiveness of each promoter region to control expression of a reporter gene in transgenic seed tissue. Characterization also includes determining the effectiveness of fragments or modifications to a promoter region to control expression of a reporter gene in transgenic seed tissue. Additional seed-specific promoter regions are provided by identifying promoter regions naturally located upstream to structural DNA sequences which are identified as homologous to the genes naturally under control of promoter regions P1, P3, P4, P6, P7, P9, P13, P14, P15, P16, P17, and P19.

Please amend the Specification at page 36, line 29 through page 37, line 5 in the following manner:

To evaluate the effectiveness of isolating and sequencing cDNAs from developing seeds to provide seed-specific novel ESTs not present in the current public data base, the entire 5' sequence data set of the present invention was compared against the *Arabidopsis* set in dbEST ~~available at www.arabidopsis.org/seqtools.html (this website is sponsored by TAIR, the *Arabidopsis* information resource which is supported mainly by NSF)~~. Of the 10,485 BLASTN results returned, 6,360 (60.9 %) showed BLASTN scores (high scoring segment pairs, HSP) of less than 50. Based on these scores it is estimated that approximately 60 % of the ESTs of the present invention are not represented in the public *Arabidopsis* EST set, and therefore many of these probably correspond to genes specifically expressed in developing seeds of *Arabidopsis*.

Please amend the Specification at page 40, line 32 through page 41, line 15 in the following manner:

To monitor seed-specific gene expressions, mRNA samples from seeds, leaves and roots of *Arabidopsis* were isolated, and reverse transcribed with oligo-dT primers into first strand cDNA fluorescent probes, as described in Example 2. The mRNA isolated from seeds was the reference to which the samples from leaves and roots were compared.

Each tissue comparison was performed at least twice, using in most cases independently isolated RNA samples as starting material. For repeated experiments, the probe pairs contained the fluorochromes Cy3 and Cy5 in opposite orientation. Results of repeated experiments were only used for further analyses if the ratios of all data points on the array showed a correlation coefficient close to one. To eliminate highly variable and therefore less reliable expression data, data was used for further analysis only if at least two experiments showed the same trend of expression. Averaging ratios across experiments was considered a less stringent strategy, because it neglects the variability between measurements (DeRisi *et al.*, Science, 278:680-686 (1997)). This is particularly true when low tissue mass (as with developing *Arabidopsis*[[_]]seeds) is a limitation for the number of feasible experiments. For the experiments described here, over 20 hours of dissection of developing seeds from siliques was required to harvest material for a single fluorescent probe.

Please amend the Specification at page 41, lines 16-28 in the following manner:

The data was analysed as a scatter plot of the data for seed vs leaf. It is clear from this representation that the majority of genes analyzed fall near the X-axis and have less than a two fold difference in signal intensity between the leaf and seed probes. Thus, although the microarray was based on a set of ESTs primarily derived from sequencing of a seed cDNA library, the overall expression pattern clearly indicates that a large proportion of seed expressed genes are also expressed in other tissues. These data support the general conclusion based on hybridization analysis of RNA complexity that 60-77% (the majority) of plant genes do not have strong tissue-specific expression (Okamura and Goldberg, 1989; Kamalay *et al.*, Cell, 19:935-946 (1980)). Expression analyses with smaller and non-seed specific arrays from *Arabidopsis*[[_]]detected comparable amounts of tissue specific (Ruan *et al.*, Plant J., 15:821-833 (1998)) or differentially expressed genes (Desprez *et al.*, Plant J., 14:643-652 (1998); Kehoe *et al.*, Trends Plant Sci., 4:38-41 (1999); Richmond *et al.*, Curr. Opin. Plant Biol., 3:108-116 (2000)).

Please amend the Specification at page 50, lines 7-21 in the following manner:

Accordingly, in some embodiments, the present invention provides methods by

which genomic sequences under control of the promoter regions of the present invention (as for example, genes 1-20 as described in Table 2, as shown in Figure 18) are used to identify additional homologous genomic sequences, preferably from other plants; the promoter regions of these homologous genomic sequences are then identified and isolated as described previously. Thus, in some aspects of the present invention, an at least partial genomic sequence of a plant is analyzed for sequences which are homologous to the *Arabidipsis* sequences which are identified as being specifically expressed in seeds (for example, those Arabidopsis sequences listed in Table 2, as shown in Figure 18). For example, BLAST searches (Altshul *et al.*, Nucleic Acids Res. 25:3389-3402 (1997); ~~http://www.ncbi.nlm.nih.gov/blast~~) may be utilized to search for nucleic acids having homology (for example, greater than 60%, 70%, 80%, or 90%) to the *Arabidopsis* sequences identified as expressed seed-specifically. Once homologous seed-specific genetic sequences are identified and isolated, they can be used to isolate promoter sequences as described above.

Please amend the Specification at page 67, line 15 through page 68, line 4 in the following manner:

For data set I, sequences were exported to plain (ASCII) text files that were used for similarity searches against GenBank using BLASTX version 1.4.11 (Altschul *et al.*, J. Mol. Biol., 215:403-410 (1990)). Sequences were first reformatted with the GCG (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc.) program REFORMAT, and the searches were done in batches using shell or PERL scripts that used GCG NETBLAST for each sequence. For data set II, the FASTA file produced by PHRED/PHD2FASTA was processed by PERL scripts to do BLASTX searches with default parameters. PERL scripts were used to assess the level of ambiguity in the DNA sequences (FASTA files) and estimate quality of the sequences based on the .qual files produced by PHRED/PHD2FASTA. The BLASTX searches were done over a period of 12 month from 9/2/98 to 9/21/99 using the most recent releases of GenBank. A subset was periodically retested (see below). The output from BLASTX was processed with PERL scripts to extract the top scoring hit from each result file. The following information for the top scoring entry in each result file was retained: gene identifier,

description, BLAST score, probability, percent identity, alignment length, and reading frame. These results were compiled in text files. Each result was manually interpreted and categorized according to predicted biochemical function. BLASTN searches were done against a subset of dbBEST ~~(available at http://www.arabidopsis.org/seqtools.html)~~ containing only Arabidopsis sequences using a FASTA file with all raw sequences. Standalone BLASTN version 2.0.9 running under linux 5.2 was used for this analysis.

Please amend the Specification at page 68, line 21 to page 69, line 17 in the following manner:

The plasmids of 2715 selected cDNA clones were collected from data set I. The inserts of the cDNAs were amplified by PCR in a 96-well format using primer pairs specific for the vector ends (for inserts in pBluescript SK-: T7, 5′-GTAATACGACTCACTATAGGGC (SEQ ID NO: 13), and 5′ extended M13 reverse, 5′-ACAGGAAACAGCTATGACCATG (SEQ ID NO: 14); for inserts in pZipLox1: M13 forward, 5′-CCCAGTCACGACGTTGTAAAACG (SEQ ID NO: 15) and M13 reverse, 5′-AGCGGATAACAATTTCACACAGG) (SEQ ID NO: 16). PCR reactions of 100 μL volume contained 0.4 μM of each primer, 0.2 μM of each desoxynucleotide, 10 mM Tris, 50 mM KCl, 3.0 mM $MgCl_2$, 3 U *Taq* DNA polymerase (Promega, Madison) and ~10 ng plasmid template. The reactions were run on a Perkin Elmer 9700 Thermoblock using an amplification program of 3 min denaturation at 94 C, 5 precycles of 30 s at 94 C, 30 s at 64 C, 2 min at 72 C, followed by 30 cycles of 30 s at 94 C, 30 s at 60 C, 2 min at 72 C and terminated by 7 min extension at 72 C. The PCR products were precipitated by adding 200 μL ethanol (95%) and 10 μL sodium acetate (3M, pH 5.2) and centrifugation at 3200 g and 4 C for 60 min. After washing with 80% ethanol, the DNA was resuspended in 20 μL 3x SSC. The yield and purity of the PCR products was analyzed by agarose gel electrophoresis. PCR samples showing by agarose gel analysis concentrations less than 0.2 μg/μL and/or double bands were repeated. If possible, alternative clones from the cDNA clone collection were used to repeat the PCR experiments. To reduce the cross-contamination risk in the 96-well format, failed PCRs were not removed from the sample set, and as a result the number of PCR samples for printing increased by approximately 20%.

Please amend the Specification at page 72, line 28 through page 73, line 21 in the following manner:

Hybridized microarrays were scanned sequentially for Cy3 and Cy5 labeled probes with a ScanArray® 3000 laser scanner at a resolution of 10 μm. In order to maximize the dynamic range of each scan without saturating the photomultiplier tube and to balance the signal intensities of the two channels approximately, laser power and PMT settings of the instrument were adjusted according to the Auto-Range and Auto-Balance features of the instrument. Signal quantitation was performed with the ScanAlyze 2.21 software written by Michael Eisen ~~(available on the Internet: http://rana.stanford.edu/software)~~. The two intensity values of duplicated DNA spots were averaged and used to calculate the intensity ratios between the two channels. Ratios below 1.0 were inverted and multiplied by -1 to aid their interpretation. Intensity values below three times their local background were deemed non-significant and excluded from further data analysis. Since subtraction of the local background from the intensity values often results in artificially high ratios, this operation was not performed for calculating the ratios. Normalization of the intensity values from the two channels was performed by stepwise exclusions of 5% of the highest and 5% of the lowest ratios and calculating for the remaining subsets the mean ratios. Usually, after excluding 15% of the highest and 15% of the lowest values, the calculated mean ratios reached a plateau, which showed only minor changes in the smaller subsets. The average value of the remaining 70% ratios was used to normalize the intensity ratios as close to 1.0 as possible. The accuracy of this filter method was evaluated by comparing it with the normalization factor calculated from the intensity ratios of the human mRNAs spiked into the labeling reaction. In general, the two methods resulted in relatively similar normalization factors. However, since external RNA controls disregard purity and integrity problems of the actual RNA samples, their use for normalization is more error prone than the filter method used for this study.

Please amend the Specification at page 74, lines 2-9 in the following manner:

Individual EST sequences were compared using BLAST against *Arabidopsis* genomic sequences larger than 10 Kb using the TAIR server manually ~~(www.arabidopsis.org/blast/)~~. After the positions of these EST sequences in the genome were determined, approximately 20 Kb flanking sequences of 30 genes were analyzed by Gene Identification Programs such as GenScan, GeneFinder and NetStart to determine the positions of ATG translation starts. The promoter regions were defined as those

regions approximately 1 Kb upstream of ATG; these regions were then selected for PCR amplification.

Please amend the Specification at page 74, lines 19-28, in the following manner:

Control vectors contained a GUS expression vector with either a napin or phaseolin promoter. For example, the promoter region of the napin (napA) gene in Brassica napus was amplified by using a forward primer CG aagctt TCTTCATCGGTGATT (SEQ ID NO: 17) and reverse primer GGTCG gaattc GTGTATGTTTT (SEQ ID NO: 18). The PCR product was digested by Hind III and EcoR I, then inserted into SK+ vector and confirmed by sequencing. The napin promoter was cut by Hind III and BamH I and inserted into a GUS expression vector such that GUS is under control of the napin promoter region. In a similar fashion, a GUS expression vector under control of a phaseolin promoter region was constructed; the phaseolin promoter region is described in patent US 5,504,200.

Please amend the Specification at page 76, line 7 in the following manner:

[[~~I~~]]G. Identification of Effective Promoter Regions

Please amend the Specification at page 77, line 7 in the following manner:

[[~~J~~]]H. Tissue-specificity of expression pattern of different promoter regions

Please amend the Specification at page 77, lines 8-17 in the following manner:

The pattern of expression of the six best seed promoter constructs was examined by plants transformed with the promoter-GUS constructs, where the promoters were one of the six promoters P1, P3, P4, P6, P16 and P17, or a napin or phaseolin promoter. The localization of by GUS expression sites in the transgenic plants was determined ~~by~~ GUS histochemical staining of young seedlings, roots, primordial tissue, floral tissue, vascular tissue, maturing leaves, and siliques. All of the transgenic plants had GUS activities in the cotyledon and hypocotyl of young seedlings. This is thought to be caused by the residue of GUS in the seed. Interestingly, promoter P4 results in GUS activity in the

floral tissues and young siliques (Figure 21). In addition, GUS activity was detected in the anther and pollen tissues of plants transformed with the phaseolin construct.

Please amend the Specification at page 77, line 18 in the following manner:

[[~~K~~]]**I.** **Timing of GUS expression during seed development.**

Please amend the Specification at page 77, lines 19-26 in the following manner:

The timing of expression of the candidate promoters at different embryo stages was also examined. Embryos were collected for GUS histochemical staining analysis at 4, 5, 6 , 7, 8, 9, 10, 12, 14 DAF. The GUS activities from the promoters ~~promtores~~ P6 and P16 were not high enough for reliable observation, so they were excluded from this analysis. The GUS expression profiles are shown in Figure 22. These expression profiles show that all the promoters start to express in mid or mid-late embryo stage. Moreover, the napin, P3, and P17 promoters result in GUS expression about 1 or 2 days earlier than do the phaseolin, P1, and P4 promoters.

Please amend the Specification at page 78, line 1 in the following manner:

[[~~L~~]]**J.** **Effects of copy number and chromosomal position on promoter activity**